

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-239799

(43)Date of publication of application : 12.09.1995

(51)Int.Cl. G06F 11/20
G06F 3/06
G06F 13/00

(21)Application number : 06-301109

(71)Applicant : INTERNATL BUSINESS MACH
CORP <IBM>

(22)Date of filing : 05.12.1994

(72)Inventor : KERN ROBERT F
MICKA WILLIAM F
MIKKELSEN CLAUS W
PAULSEN MICHAEL A
SHOMLER ROBERT WESLEY

(30)Priority

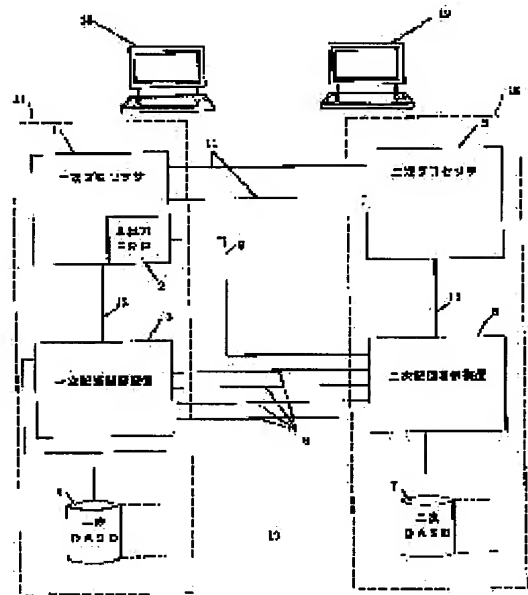
Priority number : 94 199448 Priority date : 22.02.1994 Priority country : US

(54) METHOD FOR PROVIDING REMOTE DATA SHADOWING AND REMOTE DATA DUPLEX SYSTEM

(57)Abstract:

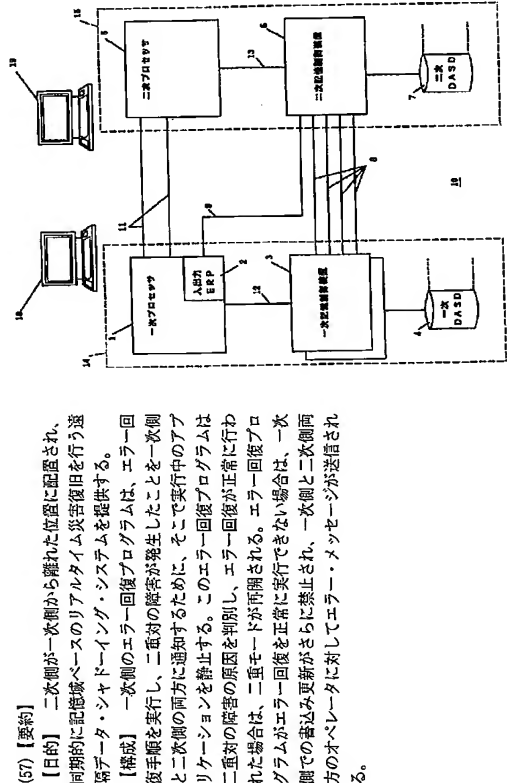
PURPOSE: To provide a remote data shadowing system which synchronously performs a real-time disaster recovery in a memory location base with a secondary side placed at a separated location from a primary one.

CONSTITUTION: An error recovery program on a primary side 14 executes error recovery procedures and then stops the application in its execution mode to notify primary and secondary sides 14 and 15 that dual pair of faults take place. The error recovery program discriminates the cause of the dual pair of faults and restarts a dual mode if an error recovery is normally performed. When the error recovery program can not normally execute the error recovery, the update of writing on the primary side is further inhibited and an error message is sent to the operators of both primary and secondary sides 14 and 15.



| (5) Int.Cl. ⁴ | G 0 6 F | 1 1/20 | 3 1 0 A | 3 0 4 E | 3 0 1 P | 特許庁 | 特許庁 |
|--------------------------|---|--------|---------|---------|---------|-----|-----|
| (5) Int.Cl. ⁴ | G 0 6 F | 1 1/20 | 3 1 0 A | 3 0 4 E | 3 0 1 P | 特許庁 | 特許庁 |
| (2) 出願番号 | 特開平6-301109 | | | | | | |
| (22) 出願日 | 平成6年(1994)12月5日 | | | | | | |
| (31) 優先権主張番号 | 1 9 9 4 4 8 | | | | | | |
| (32) 優先日 | 1994年2月22日 | | | | | | |
| (33) 優先権主張国 | 米国 (US) | | | | | | |
| (71) 出願人 | 390009531 | | | | | | |
| | インクナショナル・ビジネス・マシーンズ・コーポレーション | | | | | | |
| | INTERNATIONAL BUSINESS MACHINES CORPORATION | | | | | | |
| | アメリカ合衆国10904、ニューヨーク州アーモンク (所在地なし) | | | | | | |
| (72) 発明者 | ロバート・フレデリック・カーン | | | | | | |
| | アメリカ合衆国85730 アリゾナ州ツーソン イースト・カレッジ・ストリート8338 | | | | | | |
| (74) 代理人 | 弁理士 合田 義 (外2名) | | | | | | |
| | 最終頁に続く | | | | | | |

(54) [発明の名称] 遠隔データ・シャドーイングを提供する方法および遠隔データ二重化システム



(2) 特開平7-239799

一、メッセージを送るためのステップ (1) をさらに含むことを特徴とする、請求項6に記載の遠隔データ・シャドーイングを提供する方法。

【請求項8】 ステップ (1) によって、一次記憶サブシステムへの後続の入出力書込みがさらに防止されることを特徴とする、請求項7に記載の遠隔データ・シャドーイングを提供する方法。

【請求項9】 一次プロセッサと二次プロセッサとの間でエラー回復のための通信が行われることを特徴とする、請求項5に記載の遠隔データ・シャドーイングを提供する方法。

【請求項10】 一次記憶サブシステムと二次記憶サブシステムとの間でエラー回復のための通信が行われることを特徴とする、請求項5に記載の遠隔データ・シャドーイングを提供する方法。

【請求項11】 一次側と二次側を有し、データ・バックアップのために二次側が一次側から書込み更新を受け取り、一次側を使用不能にする災害が発生した場合に、二次側がリアルタイムのデータ可用性を提供できるように、二次側が一次側から十分離れた位置に配置される遠隔データ二重化システムにおいて、遠隔データ二重化システムが、

書込み入出力更新を生成する少なくとも1つのアプリケーションを実行する、一次側の一次プロセッサと、書込み入出力更新を受け取って格納するための一次記憶サブシステムと、

データ・リンクによって一次プロセッサに接続される、一次側の二次プロセッサと、

一次プロセッサに接続され、通信リンクを介してさらに一次記憶サブシステムに接続され、書込み入出力更新のバックアップのために記憶域ベースの二重対を提供し、受け取ったときの特定の同期順に書込み入出力更新が書き込まれる、二次側の二次記憶サブシステムと、

一次プロセッサ内にあって、二次記憶サブシステムに接続され、二重対の障害が発生した場合にエラー回復を行うためのエラー回復手段とを含む遠隔データ二重化システム。

【請求項12】 エラー回復手段が、一次プロセッサと二次側の両方に二重対障害発生モードを報告することを特徴とする、請求項11に記載の遠隔データ二重化システム。

【請求項13】 エラー回復手段が、二重対の障害の原因を判断しながら、一次プロセッサで実行中のアプリケーションを停止することを特徴とする、請求項11に記載の遠隔データ二重化システム。

【請求項14】 エラー回復手段が、一次記憶サブシステムにセンス入出力を出すことを特徴とする、請求項13に記載の遠隔データ二重化システム。

【請求項15】 エラー回復手段がエラー回復を完了し、二重モードが再確立された後、エラー回復手段が一次プロセッサに二重モードを報告する方法。

1

【特許請求の範囲】

【請求項1】 一次側が一次記憶サブシステムに接続された一次プロセッサを含み、前記一次プロセッサが一次記憶サブシステムへの入出力書込み更新の原因となる少なくとも1つのアプリケーションをそこで実行し、前記一次プロセッサがそこで実行される入出力エラー回復プログラム (A出力ERP) をさらに有し、二次側が二次記憶サブシステムに接続された二次プロセッサを含み、前記二次側が入出力書込み更新を同期的にシャドーイングし、前記二次記憶サブシステムが入出力ERPと通信する、遠隔データ・シャドーイングを提供する方法において、前記方法が、

(a) 前記一次記憶サブシステムから前記二次記憶サブシステムに入出力書込み更新を送るステップと、

(b) 前記一次および二次記憶サブシステム間に二重対が確立されているかどうかを判断するステップと、

(c) 一次記憶サブシステムから一次プロセッサに障害が発生した二重対を報告するステップと、

(d) 二重対の障害が発生した場合に、少なくとも1つのアプリケーションから入出力ERPに一次プロセッサの制御権を移行するステップと、

(e) 二重対の障害の原因を判断するステップと、

(f) 入出力ERPから二次記憶サブシステムに二重対障害発生メッセージを送るステップとを含む方法。

【請求項2】 二重対の状況が設定されるか回復された場合に前記一次記憶サブシステムが更新の書込みを続行することを特徴とする、請求項1に記載の遠隔データ・シャドーイングを提供する方法。

【請求項3】 障害が発生した二重対を報告する際に前記一次記憶サブシステムが前記一次プロセッサに入出力状況としてチャネル終了/装置終了/装置チェックを報告することがステップ (c) に含まれることを特徴とする、請求項2に記載の遠隔データ・シャドーイングを提供する方法。

【請求項4】 入出力ERPが一次記憶サブシステムにセンス入出力を出すことがステップ (d) に含まれることを特徴とする、請求項3に記載の遠隔データ・シャドーイングを提供する方法。

【請求項5】 二重対の障害が発生した場合に入出力ERPによるエラー回復を実行し、二次側でデータ安全性を維持するためのステップ (g) をさらに含むことを特徴とする、請求項4に記載の遠隔データ・シャドーイングを提供する方法。

【請求項6】 エラー回復が正常に完了した後に入出力ERPから少なくとも1つのアプリケーションに制御権を返すためのステップ (h) をさらに含むことを特徴とする、請求項5に記載の遠隔データ・シャドーイングを提供する方法。

【請求項7】 エラー回復が正常に完了できなかった場合に入出力ERPを介して二次側に障害発生同期遠隔コピー

2

一、メッセージを送るためのステップ (1) をさらに含むことを特徴とする、請求項6に記載の遠隔データ・シャドーイングを提供する方法。

【請求項8】 ステップ (1) によって、一次記憶サブシステムへの後続の入出力書込みがさらに防止されることを特徴とする、請求項7に記載の遠隔データ・シャドーイングを提供する方法。

【請求項9】 一次プロセッサと二次プロセッサとの間でエラー回復のための通信が行われることを特徴とする、請求項5に記載の遠隔データ・シャドーイングを提供する方法。

【請求項10】 一次記憶サブシステムと二次記憶サブシステムとの間でエラー回復のための通信が行われることを特徴とする、請求項5に記載の遠隔データ・シャドーイングを提供する方法。

【請求項11】 一次側と二次側を有し、データ・バックアップのために二次側が一次側から書込み更新を受け取り、一次側を使用不能にする災害が発生した場合に、二次側がリアルタイムのデータ可用性を提供できるように、二次側が一次側から十分離れた位置に配置される遠隔データ二重化システムにおいて、遠隔データ二重化システムが、

書込み入出力更新を生成する少なくとも1つのアプリケーションを実行する、一次側の一次プロセッサと、

書込み入出力更新を受け取って格納するための一次記憶サブシステムと、

データ・リンクによって一次プロセッサに接続される、一次側の二次プロセッサと、

一次プロセッサに接続され、通信リンクを介してさらに一次記憶サブシステムに接続され、書込み入出力更新のバックアップのために記憶域ベースの二重対を提供し、受け取ったときの特定の同期順に書込み入出力更新が書き込まれる、二次側の二次記憶サブシステムと、

一次プロセッサ内にあって、二次記憶サブシステムに接続され、二重対の障害が発生した場合にエラー回復を行うためのエラー回復手段とを含む遠隔データ二重化システム。

【請求項12】 エラー回復手段が、一次プロセッサと二次側の両方に二重対障害発生モードを報告することを特徴とする、請求項11に記載の遠隔データ二重化システム。

【請求項13】 エラー回復手段が、二重対の障害の原因を判断しながら、一次プロセッサで実行中のアプリケーションを停止することを特徴とする、請求項11に記載の遠隔データ二重化システム。

【請求項14】 エラー回復手段が、一次記憶サブシステムにセンス入出力を出すことを特徴とする、請求項13に記載の遠隔データ二重化システム。

【請求項15】 エラー回復手段がエラー回復を完了し、二重モードが再確立された後、エラー回復手段が一次プロセッサに二重モードを報告する方法。

50

ロセッサで実行中のアプリケーションに制御権を返すことを特徴とする、請求項14に記載の遠隔データ二重化システム。

【請求項16】エラー回復手段がエラー回復を正常に完了できない場合に、エラー回復手段が一次記憶サブシステムの書込み出力更新をならに防止することを特徴とする、請求項15に記載の遠隔データ二重化システム。

【請求項17】一次プロセッサに接続され、エラー回復手段からエラー・メッセージを受け取るための一次コンソールと、

10

【請求項18】一次および二次記憶サブシステムがそれぞれ複数の記憶制御装置を含むことを特徴とする、請求項11に記載の遠隔データ二重化システム。

【請求項19】通信リンクが、複数のエンタープライズシステム接続（ESCON）リンクを含むことを特徴とする、請求項18に記載の遠隔データ二重化システム。

【請求項20】一次側と二次側を有し、データ・シャドリングのために二次側が一次側から書込み出力更新を同期的に受け取り、一次側が使用不能になった場合に、二次側がリアルタイムのデータ可用性を提供できるように、二次側が一次側から十分離れた位置に配置され、記憶域ベースのリアルタイム災害復旧機能を提供する遠隔データ二重化システムにおいて、該遠隔データ二重化システムが、

書込み出力更新を生成する複数のアプリケーションを実行する、一次側の一次プロセッサと、

書込み出力更新を受け取るための複数の一次記憶制御装置と、

書込み出力更新を格納するための複数の一次記憶装置と、

データ・リンクによって一次プロセッサに接続される、二次側の二次プロセッサと、

二次プロセッサに接続され、さらに複数のファイバ・リンクを介して複数の一次記憶制御装置に接続され、書込み出力更新をシャドリングするために記憶域ベースの二重対を提供し、書込み出力更新が生成された順に提供される、二次側の複数の二次記憶制御装置と、

複数の二次記憶制御装置に接続され、そこに書込み出力更新を格納するための複数の二次記憶装置と、

一次プロセッサ内にあって、複数の二次記憶制御装置に接続され、二重対の障害が発生した場合に前記複数のアプリケーションから制御権を引き継ぎ、エラー回復を行わなければならない、二重対の障害の特徴に関し、一次プロセッサおよび複数の二次記憶制御装置に二重対障害発生メッセージを出し、エラー回復が失敗に終わった場合にさら

に二重対の障害が発生した場合に、二重対の障害発生メッセージを出し、エラー回復が失敗に終わった場合にさら

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

に二重対の障害が発生した場合に、二重対の障害発生

域は、テープまたはテープとDASDドライブラを含む。このレベルの記憶域では、必要なデータ記憶媒体を選択して装填するにロボットが必要になるため、ドライブ内のデータへのアクセス速度はさらに低下する。利点は、テラバイト規模のデータ記憶など、データ記憶容量が非常に大きい割に費用が削減される点である。テープ記憶装置は、バックアップ目的で使用されることが多い。すなわち、階層に保管するのために複製される。テープまたはドライブ内に格納したデータへのアクセス速度は、現在、秒単位である。

【0005】データの損失は、企業にとって破壊的なものになる恐れがあるので、多くの企業ではバックアップのデータ・コピーを用意することが必須になっている。一次記憶レベルで消失したデータの復旧に要する時間も、復旧に関し考慮すべき重要な項目である。テープまたはドライブ内のバックアップ速度の改善策としては、追加のDASDにデータが書き込まれるように追加のDASDを使用する方法がある（ミラーリングと呼ばれる場合もある）。この場合、一次DASDで障害が発生しても、データを得るために二次DASDを頼りにすることができ、この方法の欠点は、必要なDASDの数が二倍になることである。

【0006】記憶装置を二重に設ける必要性を克服するもう1つのデータ・バックアップ方法としては、低価格装置の冗余地レイド（RAID）構成にデータを書き込む方法がある。この方法では、多くのDASDにデータを割り当てるようにデータが書き込まれる。1つのDASDで障害が発生しても、残りのデータとエラー修正手順を使用すれば、消失したデータを復旧できる。現在では、数種類のRAID構成が使用できる。

【0007】一般に、記憶装置または記憶媒体で障害が発生した場合のデータの復旧には、上記のバックアップ方法で十分である。このようなバックアップ方法では、二次データは一次データのミラーになる。つまり、二次データは一次データと同じボリューム通し番号（VOLSER）とDASDアドレスを持つ。これらの方法は、装置障害の場合だけ有効である。しかし、ミラーリングされた二次データを使用しても、システム障害の復旧は行えない。このため、地震、火災、爆発、台風など、システム全体またはシステム使用現場を破壊する災害が発生した場合にデータを復旧するためには、さらに保護が必要である。つまり、災害復旧のためには、一次データから離れた位置にデータの二次コピーを格納しておく必要がある。災害保護を行う既知の方法としては、毎日または毎週、テープにデータをバックアップする方法がある。この場合、車両でテープを収集し、通常、一次データ位置から数キロメートル離れた安全な保管区域にテープを移送する。このバックアップ計画には、バックアップ

20

【0009】したがって、最小限の制御データを使用して、一次処理側のデータと一致するデータのリアルタイム更新を行う方法および装置を提供することが望ましい。この方法および装置は、復旧される特定のアプリケーション・データから独立して、つまり、特定のアプリケーション・データをベースとするのではなく、汎用記憶媒体をベースとして機能する。

30

【0010】

【発明が解決しようとする課題】本発明の目的は、一次側で行った書込み更新を二次側にシャドリング（shadowing）するための改良された設計および方法を提供することにある。

10011

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

40

【0011】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

50

【0012】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0008】さらに最近導入されたデータ災害復旧方法としては、遠隔方式だけでなく、遠隔方式でもデータがバックアップされる。遠隔二重コピーがある。あるホスト・プロセッサから別のホスト・プロセッサへ、またはある記憶制御装置から別の記憶制御装置へ、あるいはこれらを組み合わせた方法で二重データを送信するには、このプロセスを実現するために相当な量の制御データが必要になる。しかし、オーバーヘッドが高いため、二次側が一次側の処理に遅れないようにする能力が妨げられる可能性がある。このため、災害が発生した場合に二次側が一次側を復旧する能力が得られる。

【0009】したがって、最小限の制御データを使用して、一次処理側のデータと一致するデータのリアルタイム更新を行う方法および装置を提供することが望ましい。この方法および装置は、復旧される特定のアプリケーション・データから独立して、つまり、特定のアプリケーション・データをベースとするのではなく、汎用記憶媒体をベースとして機能する。

【0010】

【発明が解決しようとする課題】本発明の目的は、一次側で行った書込み更新を二次側にシャドリング（shadowing）するための改良された設計および方法を提供することにある。

【0011】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0012】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0013】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0014】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0015】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0016】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0017】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0018】

【課題を解決するための手段】本発明の第一の実施例によれば、遠隔データ・シャドリングを提供するための方法は、一次記憶サブシステムに接続された一次プロセッサを有する一次側を含み、一次プロセッサは、一次記憶サブシステムへの入出力管理更新の原因となる少なくとも1つのアプリケーションを実行する。さらに一次プロセッサは、そこで実行される入出力エラー回復プログラム（入出力ERRP）を有する。二次側は、二次記憶サブシステムに接続された二次プロセッサを含み、二次側は書込み更新を同期的にシャドリングし、二次記憶サブシステムは入出力ERRPと通信する。本方法は、一次記憶サブシステムから二次記憶サブシステムへ書込み更新を送信すること、一次記憶サブシステムと二次記憶サブシステムとの間に二重対が確立しているかどうか

【0019】

を判断することを含む。二重対の障害は、一次記憶サブシステムから一次プロセッサに報告される。一次プロセッサは、二重対の障害が発生したときに少なくとも1つのアプリケーションから入出力ERPに制御権を移転し、入出力ERPは、二重対の障害の原因を判別し、二重対障害発生メッセージを二次記憶サブシステムに送信する。

【0012】本発明の他の実施例では、遠隔データ二重化システムは一次側と二次側を混合し、二次側は、データバックアップのために一次側から書き込み更新を受け取る。二次側は、一次側を使用不能にする災害が発生した場合には二次側がリアルタイム、データ可用性を提供できるように一次側から十分離れた位置に配置される。一次側の一次プロセッサは、書き込み出力更新の原因となる少なくとも1つのアプリケーションを実行し、一次記憶サブシステムは、書き込み出力更新を受け取って格納するためにそれに接続されている。二次プロセッサは、二次側に配置され、データ・リンクによって一次プロセッサに接続される。同じ二次側に位置する二次記憶サブシステムは、二次プロセッサに接続され、さらに通信リンクを介して一次記憶サブシステムに接続され、二次記憶サブシステムは、受け取った書き込み出力更新が二次記憶サブシステムに書き込まれるような、書き込み出力更新のバックアップ用の記憶領域の二重化を提供する。一次プロセッサの出力からデータ・リターン回復は、二重化の障害が発生した場合にデータ・リターン回復を行うために、二次記憶サブシステムに接続される。

【0013】本発明の上記およびその他の目的、特徴、および利点は、添付図面に図示する本発明の実施例に関する以下の詳細な説明から明らかになるだろう。

【0014】
【実施例】 一般的なデータ処理システムは、データを計
算処理して、データ機能記憶装置サブシステム/多重
仮想記憶システム (DFSMS/MVS) ソフトウェア
などを実行するために、IBMシステム/360または
IBMシステム/370プロセッサなどのホスト・プロ
セッサの形態をとり、少なくとも1台のIBM 399
0記憶制御装置がそれに接続され、その記憶制御装置
が、メモリ制御装置と、それに組み込まれた1つまたは
複数のタイプのキャッシュ・メモリを含む場合があ
る。さらに記憶制御装置は、IBM 3380または3
390 DASDなどの1群の直接アクセス記憶装置
(DASD) に接続されている。ホスト・プロセッサが
実質的な計算能力を提供するのに対し、記憶制御装置
は、大規模データベースを効率よく搬送し、ステージ
(stage) / デスステージ (destage) し、変換し、全般的
にアプリケーションに必要なる機能を提供する。

【0015】一般的なデータ処理システムの災害復旧保護では、一次DASDに格納した一次データを二次側または遠隔地でバックアップする必要がある。一次側と二

二枚側との距離は、ユーザが受け入れられる危険のレベルによって決まり、緩急ローメータトルンから数千キロメートルの範囲が可能である。二枚側または遠隔地は、バックアップ・データ・コピーを提供するだけでなく、一時的なシステムが使用不能になった場合に一次システムの処理を引き継ぐのに十分なシステム情報を持つていなければならぬ。その理由は、主に、単一の記憶制御装置では一次側と二枚側にデータを書き込めないためである。むしろ、二枚側の両方にデータを書き込む二枚DASDストリングに一次記憶制御装置に接続された一次DASDストリングに一次データが格納されるのに対し、二次記憶制御装置に接続された二次DASDストリングには二次データが格納されるのである。

【0016】二次側は、一次側から十分離れている必要があるだけでなく、一次データをリアルタイムでバックアップできないわけではない。二次側は、一次データが更新されたときに、最小限の遅延で一次データをバックアップする必要がある。しかも、二次側は、一次側で実行され、データまたは更新を生成するアプリケーション・プログラム（たとえば、IMSやDDB2）を考慮せずに、一次データをバックアップしなければならない。二次側に要求される難しい課題は、二次データの順序が整合していないなければならないことである。つまり、二次データは一次データと同じ順序でコピーしなければならない（つまり（順序整合性）、順序整合性）についてかなり考慮を要する問題である。順序整合性は、それぞれが1つのデータ処理システム内の複数のDASDを制御する共通制御装置が複数存在するためにさらに複雑になっている。順序整合性がなくとも、一次データと一致しない一次データが生成され、その結果、災害復旧が破壊される恐れがある。

【0017】遠隔データの二重化は、同期と非同期の2つの一般的なカテゴリに分けられる。同期遠隔コピーでは、データを一側を二側に通し、一次DASDの出力操作を終了する（一次ホストにチャネル終了（CE）装置終了（DE）を出力する）前にこのようなデータの受け取りを確認する必要がある。このため、同期コピーは、二側側の確認を待っている間に一次DASDの出力応答時間が遅くなる。一次側の出力応答延ばし（これは遠隔距離とデータの距離に比例して長くなる）は、同期コピーは、システム・オーバヘッドを比較的小さくして、順序が整合したデータを二側側に提供する。

【0018】非同期遠隔コピーでは、二次側でデータが雄読される前に一次DASDの入出力操作が完了する（一次カストにチャネル終了（CE）と装置終了（DE）を出力するため、一次側のアプリケーション・システムのパフォーマンスが向上する。このため、一次DASDの入出力応答時間は二次側までの距離に依存せ

ず、一次側から数千キロメートル離れた遠隔地に二次側を設けることもできる。しかし、二次側で受け取ったデータの順서가一次側の更新順序と一致しない場合には、多くのデータ、データの順序整合性を確保するのに必要なシステム・オーバーヘッドが増加する。したがって、一次側で競争が発生すると、一次側と二次側との間で伝送中のデータが一部紛失する恐れがある。

【0019】同期データ・シャドウイング
災害復旧のための同期リアルタイム遠隔コピーでは、コピーされた複製のDASDボリュームが1つのセットを形成する必要がある。さらに、このようなセットを形成するには、各セットを構成するこれらのボリュームとそれに対応する一次側の同等物（VOLSER）を識別するために十分な量のシステム情報（VOLSER）を識別する必要がある。重要なことは、二次側が一次側と二重対（duplex pair）を形成するのは、2つまたは複数のボリュームがこのセットと同期していない、つまり、「二重対の故障」が発生したときに二次側がそれを認識しなければならぬ点である。代替案が再試行する間に二次DASDの出力が遅延せず、非同期遠隔コピーより同期遠隔コピーの方が、故障影響を軽減するやうい

一次側は、二次側用の更新を待ち行列に入れる間、一次側が執行できるようにコピーを中止または中断することができ、このように二次側が同期していないことをいつまでも災害復旧に使用できるようにするため、二次側が一側と同期しなくなる原因にならずな例外条件を認識して、二次側が必要である。一側条件や復旧処置によって、二次側が一次側と一致しなくなっている。

【0020】しかし、二次DASDが存在しないアークセクシ
ンである状態で、二次側と一次側の接続を維持しても、
内容の同期は確保されない。いつかの理由から、二次
側は一次側との同期性を失う場合がある。二次側が形成
されたときに二次側は当初同期しておらず、初期データ
・コピーが完了したときに同期し直す。一次側が二次
側に更新済みデータを書き込みない場合、一次側は二重
写を解除することもある。この場合、二次側は、更新ア
プリケーションが実行できるように、二次側が中断され
た状態で一次DASDに更新内容を書き込む。このため、つ
まり、二重対が短元されるまで、一次側は競合状態、つ

まり、現行の災害保護コピーを使わずに実行を続ける。二重効が復元されたも、二次側が直ちに同期状態になるわけではない。この時点で保護状態の更新を適用した後に、二次側は同期状態に戻る。一次側は、該当ポリシーに関する中止コマンドを一次DASDに出すことによって、二次側の同期を喪失させることもできる。この中止コマンドが終了し、二重効が再確立され、保留状態の更新がコピーされると、二次側は一次側と再同期する。また、オンライン保守のもと、同期を喪失させることができる。

【0021】二次ボリュームが一次ボリュームと同期化していない場合、二次ボリュームの再同期は、二次システムの復旧と一次側アプリケーションの再開に使用することができ、二次側の非同期ボリュームは非同期ボリュームとしない。二次側の非同期ボリュームは非同期引継ぎ手順として識別されない限りならず、二次側の復旧引継ぎ手順には、アプリケーション・アクセスを否定する（そのボリュームを強制的にオンラインにするか、そのVOLSE Rを交更する）ために非同期ボリュームを識別しなければならない。一次側ボリュームがアクセス不能になった場合に二次側を復旧するためには、二次側が呼び出されることである。このため、二次側では、すべてのボリュームの同期状態に因するすべての関連情報が必要である。二次記憶サブシステム、二次記憶制御装置とDASDは、一次側で検出された例外によって一次側が同期を解する原因となるすべての条件を判断できるわけではない。たとえば、二次側が把握していない一入出力経路またはリンクの障害のために一次側が二次側の同等機能にアクセスできない場合、一次側は二重対を解除することに対し、二次側は二重対が解除されたことを示す。

【0022】非同期二重対称リユームが存在すること
を、外部通信によって二次に通知することができる。
これは、ユーザー・システム管理機能を使用すること
でできる。一次に入力操作はチャネル終了/装置終
了/装置チェック (C/E/D/E/UC) 状態で終了し、
センス・データがエタがエタの特徴を示す。このよう
な形式の入出力構成の場合、エラー回復プログラム
(ERP)のエラーを処理し、入出力の完了を一次に
アプリケーション・ジョブに通知する前に、二次に
メッセージを送る。この場合、ERPの二重対称
メッセージを認識し、その情報を二次に確保する
のは、ユーザーの責任である。一次に代わりの
動作可能になるよう二次に依頼にされる場合は、
強制手順によって二次にDASDが二重ホストに
オンライン接続のために非同期対称リユームが
オンライン接続されていることを確認するために、
二次にDASDサブシステムに格納された同期状
況をすべてのERP二重対称メッセージと組み合
わせられる。二次に、二次に全体を示すピクチャ
が得られる。

【0023】ここで図1を参照して説明すると、同図に示されるのは、一次側14と二次側15を有し、災害復旧システム1が実行されている。一次側14は、災害復旧システム1から20キロメートル離れている。二次側15は、災害復旧システム1のアプリケーションと、システム入力および出力を介して接続されている。二次側15は、以下、入出力ERP2というプログラム2（以下、入出力ERP2）を含む。入出力ERP2は、一次側14にデータを転送する。一次側14は、DF SMS/MV Sオペレーティング・ソフトウェアを実行するIBMエンタープライズ・システム9000（ES/9000）などを含む。

との間、または最悪の場合は、一次プロセッサ1と二次プロセッサ5との間のデータバス安全性を維持するために、新たな対応等通信同期エラー回復を実行することができ

る。

【0026】図2および図3を参照して説明すると、同図にはエラー回復手順が示されている。図2のステップ201は、一次プロセッサ1で実行されるアプリケーション・プログラムが一次記憶制御装置3にデータ更新を送信することを含む。ステップ203では、そのデータ更新が一次DASD4に書き込まれ、そのデータ更新が二次記憶制御装置6にシャド・イングされる。ステップ205では、二重対の状態がチェックされ、一次側と二次側が同期しているかどうかが判別される。二重対の状態が同期状態になっている場合、ステップ207でデータ更新が二次DASD7に書き込まれ、一次プロセッサ1での処理は、そこで実行されるアプリケーション・プログラムを介して続行される。

【0027】二重対が「障害発生」状態になっている場合、ステップ209で一次記憶制御装置3は、二重対で中断または障害が発生していることを一次プロセッサ1に通知する。二重対は、通信リンク8による一次記憶制御装置3と二次記憶制御装置6との通信障害によって「障害発生」状態になる場合がある。あるいは、二重対は、一次サブシステムまたは二次サブシステムいずれかのエラーによって「障害発生」状態になる場合もある。

障害が通信リンク8で発生している場合、一次記憶制御装置3は、二次記憶制御装置6に直接、障害を連絡することができない。そこで、一次記憶制御装置3は、ステップ211で出力状況としてCE/DE/UCを一次プロセッサ1に返す。入出力ERP2は、アプリケーション・プログラムを静止させ、書き込み出力操作を要求するアプリケーションに制御権を返す前に、エラー回復とデータ安全性のためにステップ213で一次プロセッサ1の制御権を引き継ぐ。

【0028】図3は、入出力ERP2が実行する諸ステップを表している。ステップ221で入出力ERP2は一次記憶制御装置3にセンス入出力を出す。センス入出力操作は、入出力エタラの原因を記録する情報を返す。すなわち、このデータ記録情報は、具体的なエラーに関する記憶制御装置または二重対の操作に固有のものになって記憶制御装置3と二次記憶制御装置6との間の一次記憶制御装置3と二次記憶制御装置6との間の対等通信リンク8で障害が発生したことがデータ記録情報によって示された場合、ステップ223で入出力ERP2は、一次記憶制御装置3および二次記憶制御装置6に対して、関係係保プログラムを障害発生同期遠隔コピー状態に入れるように指示する記憶制御装置レベル出力操作を出す。この二次記憶制御装置6は、複数のESCOリンク9または二重対間通信リンク11を介して入出力ERP2または関係係保プログラムの状態を受け取ることができる。その結果、二重対操作の現在の状況は、一次

ロセッサ1で実行されるアプリケーションとともに、一次プロセッサ1および二次プロセッサ5の両方で維持される。コンソール18および19は、それぞれ一次プロセッサ1および二次プロセッサ5からの情報をやりとりするため設けられ、入出力ERP4は、両方のコンソール18および19に状況情報を通知する。

【0029】一次記憶制御装置3および二次記憶制御装置6への障害発生同期遠隔コピー入力操作が正常に完了すると、ステップ225でデータバス安全性が維持されていく。このため、二次側15で復旧を試みると、二次記憶制御装置6は、「障害発生同期遠隔コピー」というマークを付けたプログラムを、データ回復手段（プログラム1のそのデータの状態を判別するための従来のデータベース・ログまたはジャーナル）によってそのプログラムのデータとその同期グループ内の他のデータとの同期が取られるまで使用できないものとして識別する。

【0030】ステップ227では、障害発生同期遠隔コピーの状況更新について一次記憶制御装置3と二次記憶制御装置6で行われた入出力操作の正常終了を入出力ERP2が受け取ったかどうかを判別するテストが行われる。正常終了すると、入出力ERP2は、ステップ229で一次プロセッサ1に制御権を返す。正常終了していない場合は、ステップ231で次のレベルの復旧通知が行われる。この通知には、障害発生プログラムと、一次記憶制御装置3または二次記憶制御装置6のいずれかのそのプログラムの状況が正しくない可能性があることを、コンソール18を介してオペレータに通知することが含まれる。この通知は、そこで具体的なプログラム状況を示すために、コンソール19または共用DASDデータ・セットを介して次側15にシャド・イングされる。

【0031】ステップ233で、エラー・ログ記録データ・セットが更新される。この更新は、一次DASD4または他の記憶場所のいずれかに書き込まれ、二次側15にシャド・イングされる。このエラー回復処理が完了すると、入出力ERP2はステップ235で、障害発生書き込み出力操作に関する「永続エラー」回復を一次側アプリケーションに実行させるために、一次側アプリケーションの書き込み出力操作に「永続エラー」を通知する。エラーが修正されると、プログラム状態は、まず保留状態（変更データの再コピー）に回復し、次に全二重対に回復することができ。その後、二重対が再確立されると、データを二次DASD7に再適用することができ

る。

【0032】二重対を確立する場合、顧客の要求に応じて、プログラムをCRITICALと識別することができ。CRITICALプログラムの場合、ある操作の結果、二重対の障害が発生すると、実際のエラー箇所とは無関係に、一次プログラムの永続エラー障害が報告される。CRIT=Yの場合、障害発生時の一次DASD

406に書き込むとするその後のすべての試みは、永続エラーを受け取ることになり、対応をなす二次プログラムにシャド・イングできないデータは、その一次プログラムに一切書き込まれなくなる。このため、必要であれば、一次側アプリケーションの処理および入出力データ操作との完全同期が可能になる。

【0033】その結果、本明細書に記載する災害復旧システム10では、入出力命令（チャネル・コマンド・ワード（CCW））を有する一次ホスト処理エタラ回復手順によって、二重対から障害発生二重対へ一次および二次同期遠隔コピー・プログラムを複製して書き込むよう外部同期遠隔コピーを取り入れ、それにより、複製データの完全性を維持する。アプリケーション・ベースのパックアップではなく、データ更新がリアルタイムで書き込まれる記憶媒体ベースのバックアップが設けられている。

また、災害復旧システム10は、(1)一次および二次記憶制御装置プログラム状況更新、(2)オペレータ・メッセージまたはエラー・ログ記録共通データ・セットを介して具体的なプログラム更新通知に一次および二次ホスト・プロセスが通知すること、および(3)CRITICALプログラム識別などの、複製レベルの一次/二次状況更新を試み、プログラムが障害発生二重対になる場合は、一次プログラムへのその後の更新を防止することができる。このため、リアルタイムの完全エラー災害復旧が達成される。

【0034】非同期データ・シャド・イング非同期遠隔データ・シャド・イングは、1回の災害で一次側と二次側の両方が関連してしまふ確率を低減するたに一次側と二次側の距離をさらに大きくする必要がある場合、または一次側アプリケーションのパフォーマンスへの影響を最小限に抑える必要がある場合に使用する。一次側と二次側との距離は、現在では地球全体またはそれ以上に延長できるが、複製の一次サブシステムの背後にある複製のDASDプログラムにわたる書き込み更新を複製の二次サブシステムに同期させることは、さらに複雑である。二次記憶サブシステム上でシャド・イングするために、一次データ・ムーバを介して一次記憶制御装置から二次データ・ムーバ・レベルコード書き込み更新を送信することができるが、両者間でやりとりされる制御データの量は、最小限でなければならない。同時に、複製の記憶制御装置に隠れている複製のDASDプログラムにわたる二次システム上でレコード書き込み更新の順序を正確に再構築できるものでなければならない。

【0035】図4は、一次側421と遠隔側または二次側431とを含む非同期災害復旧システム400を示している。一次側421は、DFSMS/MVSホスト・ソフトウェアを実行するIBM ES/9000などの一次プロセッサ401を含む。一次プロセッサ401

この通信リンク408は、電話（T1、T3回線）、無線、無線/電話、マイク波、衛星などの複数の適当な通信方式によって実現できる。

【0039】非同期データ・シャドーイング・システム400は、一次DASD406へのすべてのデータ書き込みの順序が保持され、二次DASD416に適用される（すべての一次記憶制御装置405にわたるデータ書き込みの順序を保持する）ように、一次記憶制御装置405から制御データを収集する機能を含む。二次側431に送られるデータおよび制御情報は、データ保全性を保持するのに一次側421の存在が不可欠になるほど、十分なものでなければならぬ。

【0040】アプリケーション402、403は、データまたはレコード更新を生成するが、このレコード更新は、一次記憶制御装置405によって収集され、PDM404によって読み取られる。それぞれの一次記憶制御装置405は、非同期遠隔データ・シャドーイング・セッションのためにそれぞれのレコード更新をグループ化し、非特定一次DASD406のREAD要求を紹介する。PDM404にこれらのレコード更新を提供する。一次記憶制御装置405からPDM404へのレコード更新の転送は、START入出力操作の回数を最小限にしながら、各一次記憶制御装置405と一次プロセッサ401との間で転送されるデータの量を最大にするように、PDM404によって制御され、最適化される。PDM404は、非特定READ間の時間間隔を変えることで、一次記憶制御装置405とこの最適化だけでなく、二次DASD416用のレコード更新の適用期間も制御することができる。

【0041】データ保全性を維持しながら、PDM404がレコード更新を収集し、そのレコード更新をSDM414に送信するに、すべての一次記憶サブシステムにおいて二次DASD416に対して行われる一次DASD406のレコードWRITEシーケンスを再構築するに十分な制御データとともに、特定の期間の間、適切な数の時間間隔でレコード更新を送信する必要がある。一次DASD406のレコードWRITEシーケンスの再構築は、自己記述レコードをPDM404からSDM414に渡すことによって達成される。SDM414は、所与の時間間隔分のレコードが紛失しているかどうか、または不完全になっているかどうかを判断するために、その自己記述レコードを検査する。

【0042】図5および図6は、接頭部ヘッダ500（図5）と、一次記憶制御装置405によって生成されたレコード・セット情報600（図6）を含む、各自己記述レコードごとにPDM404が作成するジャーナル・レコード形式を示している。各自己記述レコードは、それぞれの時間間隔の時間順に二次DASD416に適用できるように、それぞれの時間間隔ごとにさらにSDM414によってジャーナル処理される。

は、IMSおよびDBSアプリケーションなどのアプリケーション・プログラム402および403と、一次プロセッサ（PDM）404をさらに含む。一次プロセッサ401には、そこで実行されるすべてのアプリケーション（402、403）に共通の基準を提供するために、共通シスプレックス・クロック（splex cloc k）407が設けられ、すべてのシステム・クロックまたは時間値（図示せず）がシスプレックス・クロック407に同期し、すべての時間依存プロセスが相互に正しいタイミングで動作するようになっている。たとえば、一次記憶制御装置405は、単一の一次記憶制御装置406への2回の連続する書き込み入出力操作が同じタイム・スタンプ値を示さないように、複数のレコード書き込み更新時間を確実に区別するに適した解像度に同期している。シスプレックス・クロック407の解像度（正確さではない）は重要である。PDM404は、シスプレックス・クロック407に接続された状態で図示されているが、書き込み入出力操作がそこで発生するわけではない。また、一次プロセッサ401が単一の時間基準（たとえば、単一のマルチプロセスサES/9000システム）を有する場合に、シスプレックス・クロック407は不要である。

【0036】一次プロセッサ401には、IBM 3990-6型記憶制御装置などの複数の一次記憶制御装置405が光ファイバ・チャネルなどの複数のチャネルを介して接続されている。また、各一次記憶制御装置405には、IBM 3390 DASDなどの複数の一次DASD406からなる少なくとも1つのストリングが接続されている。一次記憶制御装置405と一次DASD406によって、一次記憶サブシステムが形成される。各記憶制御装置405と一次DASD406は、個別のユニットである必要はなく、両者を組み合わせて単一のドロブにしてもよい。

【0037】一次側421から数千キロメートル離れた位置に配置される二次側431は、一次側421と同様に、そこで動作する二次データ・ムーバ（SDM）414を有する二次プロセッサ411を含む。二次プロセッサ411には、当技術分野で既知の通り、光ファイバ・チャネルなどのチャネルを介して複数の二次記憶制御装置415が接続されている。記憶制御装置415には、複数の二次DASD4161つの制御情報DASD417（複数可）が接続されている。記憶制御装置415とDASD416および417によって、二次記憶サブシステムが構成される。

【0038】一次側421は、通信リンク408を介して二次側431と通信する。より具体的には、一次プロセッサ401は、仮想的通信アクセス方式（VTA M）通信リンク408などの通信プロトコルによって、二次プロセッサ411にデータと制御情報を転送する。

【0043】ここで図5を参照すると、各レコード・セットの先頭に挿入される接頭部ヘッダ500は、接頭部ヘッダ500と、各レコード・セットごとにSDM414に送信される実際の一次レコード・セット情報600との長さの合計を記述するための総データ長501を含む。操作タイム・スタンプ502は、PDM404が現在処理している操作セットの開始時間を示すタイム・スタンプである。この操作タイム・スタンプ502は、1組の一次記憶制御装置405に対してREAD RECORD SET機能を実行する際に（シスプレックス・クロック407に応じて）PDM404によって生成される。一次DASD406の書き込みの入出力時間610（図6）は、各一次記憶制御装置405のREAD RECORD SETごとに固有のものである。操作タイム・スタンプ502は、すべての記憶制御装置で共通のものである。READ RECORD SETコマンドは、PDM404によって出されるが、以下の条件のいずれかの場合に予測できる。

(1) 一次記憶制御装置405の所定のしきい値に基づく、その一次記憶制御装置のアレンジション制込み
(2) 所定の時間間隔に基づく、一次プロセッサ401のタイムレコ込み

(3) レコード・セット情報が、使用可能であるがまだ読み取られていない未解決のレコード・セットに関する追加情報を示す場合

条件(2)では、タイム間隔を使用して、低レベル活動の期間中に、該システム431がどの程度遅れて実行するかを制御する。条件(3)は、PDM404が一次記憶制御装置405の活動に遅れないようにするためにさらに活動を駆動する処理期間中に、PDM404がすべてのレコード・セットを待ち行列処理しなかった場合に発生する。

【0044】時間間隔グループ番号503は、現行レコード・セット（整合性グループのうちの所与の時間間隔グループ）についてすべての一次記憶制御装置405にわたるレコードのセット）が属する時間間隔（操作タイム・スタンプ502とレコード総取り時間507によって境界が示される）を識別するためにPDM404が出力する。グループ内順序番号504は、所与の時間間隔グループ503内の各レコード・セットごとに一次記憶制御装置405用のアプリケーションWRITE入出力の順序を（PDM404に対して）識別するためにヘッダウェアが提供するIDである。一次SSID（補助記憶領域ID）505は、各レコード・セットごとに一次記憶制御装置405の特定の一次記憶制御装置を明確に識別するものである。二次ターゲット・ポリューム506は、パフォーマンス上の考慮事項に応じて、PDM404またはSDM414のいずれかによって割り当てられる。レコード総取り時間507は、すべての一次記憶制御装置405に共通の操作タイム・スタンプを提供し、現行

間隔のレコード・セットの終了時間を示す。

【0045】次に図8を参照して説明すると、レコード・セット情報600は、一次記憶制御装置405によって生成され、PDM404によって収集される。更新固有情報601〜610は、レコード更新が行われた実際の一次DASD406を含む各レコードの一次装置ユニット・アドレス601を含む。シリング番号/ヘッダ番号（CCHH）602は、各レコード更新ごとの一次DASD406上の位置を示す。一次記憶制御装置のセッションIDである二次SSID603は、一次SSID505と同じものである。状況フラグ604は、特定のデータ・レコード620が後に続くかどうかに関する状況情報を提供する。順序番号605および630は、レコード・セット全体（PDM404に転送されたすべてのデータ）が読み取られたかどうかを示すために各レコードに番号を1つつ割り当てる。一次DASD書き込み出力タイプ606は、各レコードについて行われた書き込み操作のタイプを識別する操作模様である。この操作模様は、更新済み、フォーマット済み、部分ドロップ・レコード・フォロー、完全トラック・データ・フォロー、消去コマンド実行、または全書き込み実行を含む。検索索引数607は、最初に読み取られたレコード・セット・データ・レコード620に関する初期位置決め情報を示す。セクタ番号608は、レコードが更新されたセクタを識別する。カウント・フィールド609は、後続の特定のレコード・データ・フィールド620の数を記述する。一次DASD406の書き込み更新が行われたホスト・アプリケーション時間は、更新時間610に記録される。特定のレコード・データ620は、各レコード更新ごとのカウント/キー/データ（CKD）フィールドを提供する。最後に、順序番号630は、読み取られたレコード・セット全体がPDM404に転送されたかどうかを示すために順序番号605と比較される。

【0046】一次DASD406でレコード更新が書き込まれたのと同じ順序でSDM414がそのレコード更新をコピーできるように、ソフトウェア・グループが呼出した整合性グループで更新レコードが処理される。整合性グループを作成するために使用する情報（すべての記憶制御装置405から収集したすべてのレコード・セットにわたる）は、操作タイム・スタンプ502、時間間隔グループ番号503、グループ内順序番号504、一次制御装置SSID505、レコード総取り時間507、一次装置アドレス601、一次SSID603、および状況フラグ604を含む。1つの時間間隔グループのすべてのレコードがSDM414側で各記憶制御装置405ごとに受け取られたかどうかを判断するために使用する情報は、時間間隔グループ番号503、グループ内順序番号504、物理制御装置ID505、および一次SSID603を含む。完全復旧可能な一次DASD406のレコード更新と同様に二次DASD416上に

レコード更新を配置するのに必要な情報は、二次ターゲット・ボリューム506、CCH602、一次DAS D着込み入出力タイプ606、検索引数607、セクタ番号608、カウント609、更新時間610、および特定のレコード・データ620を含む。

【0047】図7および図8は、復旧時間・ジャーナルデータを転送する装置を記述するものである。状態データ700とマスタ・ジャーナル800をそれぞれ示す。状態データ700は、PDM404またはSDM414のいずれかが収集した構造物情報を提供し、一次記憶制御装置のセッションID (SSID) 番し、およびその制御装置のセッションIDおよび対応するセッションおよびその制御装置のセッションIDおよび対応するセッションとを結合する。このため、構造物情報は、二次ポートユーザ710または二次DSDエクステンションが二次ポートユーザ711または二次DSDエクステンションにマップされていることを追跡する。状態データ700でも単独に拡張され、部分ポートユーザ、エクステンション712 (CCC HHからCCHHまで) を示す場合、部分ポートユーザ識別コードは、ここに記述するのと同じ非同担識別コードを使用して達成できるが、完全ポートユーザの場合より細かくなる。

【0048】マスタ・ジャーナル800は、整合性グループ番号、ジャーナル・ポイントの位置、およびレポートタイム、スタンバイを含む。また、マスタ・ジャーナル800は、整合性グループにグループ化し特定のリコード更新を維持する。状態グループ700とマスタ・ジャーナル400は、災害復旧をサポートするため、一次システム810がもはや存在しないスタンダード環境で動作できなければならない。

【0049】制御項目全体が正しく書き込まれるようにするため、タイム・スタンピング制御は各マスター・ジャーナル8000の前後に置かれる。このタイム・スタンピング制御は、さらに二次データSD417に書き込まれる。制御要領は二番項目(1)および(2)を含み、次に示す例のように二方の項目が必要実行項目になる。

- (1) タイム・スタンプ制御 | 制御情報 | タイム・スタンプ制御
 (2) タイム・スタンプ制御 | 制御情報 | タイム・スタンプ制御

いかなる時点でも、(1)または(2)のいずれかの項目が現在または有効項目になるが、有効項目は前記に等しいというタイムスタンプ制御を持つ項目である。装置復旧時には、制御情報を得るために、最新のタイムスタンプを持つ有効項目を使用する。この制御情報は、状態情報(記憶制御装置、装置、および適用される整合性グループ)に関する整合情報とともに、二次記憶装置に415にどの1つのニューロ更新が適用されたかを判別するのに使用する。

【00501】整合性グループ

所定の時間間隔の間にすべての一次記憶制御装置405にわたるすべての記録レコード・セットが二回、431で1で受け取られると、SDM414は、受け取った制御情報を読み出し、レコード更新が最初在一次DSAD406上で書き込まれたのと同じ順序でそのレコード・セットをレコード更新・グループとして二回DSAD416に適用する。このため、一次制御アプリケーションの順序（データ保全）整合性はすべて二回431で維持される。このプロセスは、以下、整合性グループの形成と評価、整合性グループの形成は、次のような仮定に基づいて行われる。（A）独立しているアプリケーション書き込みが制御装置の順序命令に違反しない場合は、どのような順序でもアプリケーション書き込みを実行する。

(B) 従属の順に実行しなればならぬとは、アプリケーションは、タイマ・スケジュールは、番込み番号1から制御施修終了、装置終了まで、完了する前に従属番号2を実行することが必要ない。(C) 第二の番込みは必ず(1)遅いタイム・スラングを持つ第一の番込みと同じレコード・セットを含むグループに入る。(2) 後続のレコード・セットを含むグループに入る。

【0051】図9を参照して説明すると、図9には、格納制御装置SSD11、SSD2、およびSSD3など（記憶制御装置はいくつも含まれることができる）の例では明解にするため3つ使用する）に関する整合性グループの形成例（整合性グループは一次側42および二次側43のいずれにも形成できる）は、図9に示される。時間間隔T1、T2、およびT3は昇順に発生するものと想定する。時間間隔SSD1、タイム・スタンプ502は、記憶制御装置SSD1、SSD2、およびSSD3について設定されている。PDM404は、時間間隔T1〜T3の間に記憶制御装置SSD1、2、および3からレコード・セット・データを入力する。時間間隔T1のSSD1、2、および3に関するレコード・セットは、時間間隔グループ1であるG1（時間間隔グループ番号503）に割り当てられる。グループ内番号504は、SSD1、2、および3のそれぞれについて示され、この場合、SSD1は11:59、12:00、および12:01に3つの更新を持ち、SSD2は12:00、および12:02に2つの更新を持ち、SSD3は11:58、11:59、および12:02に3つの更新を持つ。時間間隔T2およびT3のレコード・セットは1つ列挙されているが、簡略化のため、更新時間の例は示されていない。

【0052】ここで、二次側431で受け取った制御情報およびレコード更新に基づいて、整合性グループNを生成することができる。時間間隔グループ番号1のレコ

コード更新時間間隔グループ番号2のレコード更新より更新より遅くならないようにするため、記憶制御装置SSD1の2、および3のそれぞれ最後のレコード更新時間よりも早い読取りレコード・セット時間と等しい、最小時間も設定される。この例では、最小時間は12:01になる。最小時間と等しいそれ以上の読取りコード・セット時間と有するレコード更新はすべて整合性グループN+1に属する。1つの要求する2つのレコード更新時間間隔が等しい場合、システム・グループ・クロック4の十分な解像度が与えられる可能性はほとんどないが、時間間隔グループN内の早い順序番を持つレコード更新は、整合性グループN用のそのグループとともに保管される。ここで、レコード更新は、読取りレコード・セット時間に基づいて順序づけされる。複数のレコード更新の時間が等しい場合、小さい順序番号を持つレコード更新は、大きい順序番号を持つレコード更新の前に並べられる。これに対して、複数のレコード更新のタイム・スタンプが等しいが、より若い整合性グループ更新の場合、そのコード更新の時間よりも早い整合性グループ更新がある限り、任意の順序にすることができ、

【0053】一次記憶制御装置405が、指定の時間間隔の間に記憶リコード・セットへの応募を完了しなかった場合、その一次記憶制御装置405が完了するまで、競合シナリオを形成することはできない。また、記憶制御装置405がその応募を完了しなかった場合は、未着利込のため、シナリオの未着利込ハンドラが制御権を要求取り、操作が終了する。これによって、記憶制御装置405が通常操作に実行する。競合は、入出力が完了に至り、通常操作が実行される。競

に、整合性グループの形成では、一斉強制制装置405によって予想される書き込み操作にタイム・スタンプが付けられると予想される。しかし、プログラムによっては、タイム・スタンプが付けられずに書き込みが生成されるものもある。この場合、一斉強制制装置405は、タイム・スタンプとしてそれを返す。整合性グループの形成は、データが読み取られたタイム・スタンプに基づいて、タイム・スキャンを可能にする。整合性グループの時間的にレコード更新の世界観を容易に示せないほど、タイム・スタンプを持たないレコード更新が一定の時間間隔の間に多数発生した場合、

二重ボリュームが同期していないというエラーが発生する可能性がある。

形成する。図10を参照して説明する。このプロセスは、ステップ100から始まり、一歩421が、行うべき近隣データ・シャドワイズとして、すべてのアブリーケーション出力421を使用して、ステップ101では、サブプレック・クロックの470を同期クロック(図4)として使用し、すべてのアブリーケーション入力操作にタイム・スタンプが付けられる。PDM404は、ステップ10

Aを提示したか、レコード、セット、バックアップ内にこの
M4 04から受け取らなければならないし、しかもデータ
(またはメタ)を持つすべての読取りレコーダ・セット
・バックアップがSDM4 14によって受け取られていなければならぬ。時間間隔グループが不完全な場合は、ス
テップ1110は、必要データが受け取られるまで、
一次延滞検探率405の次のレコーダ・セットの読取
りを再試行する。エラーが発生した場合は、特定の1つ
または複数の二重バリュウム対に障害が発生している可
能性がある。120は、第一の総合性グループを受け取り、ス
テップ1110では、第二の総合性グループ・ジャーナル・コ
ードとを判別する。その第一(または第二)の総合
性グループ・ジャーナル・コードとは、最も早い操作
タイム・スタンプ502、同じ操作タイム・スタンプ
0を含むコールドである。

【0058】ステップ1130は、現行整合性グループ・ジャーナル・レコードに含まれるレコードを検査し、どのレコードがその整合性グループに含まれ、どのレコードがその整合性グループから除外される（一部のレコードは除外され、次の整合性グループ・ジャーナル・レコードに含まれる）。現行整合性グループ・ジャーナル・レコードの最後は、各グループ・ジャーナル・レコードの最後のレコードは、各グループ・ジャーナル・レコードの最後のレコードの最後（一次記憶装置405ごとに最大更新時間のうちの一（小更新時間（或小時間））として平均される（つまり、各グループ・ジャーナル・レコードの最後の更新が比較され、これら（小更新時間）405のうちの最も早いものが現行整合性グループ・ジャーナル・レコードに渡る）。

【0059】 現行整合性グループ・ジャーナル・レコードに残っているこれらのレコード更新は、ステップ1140で、更新時間610とグループ内順序番号504に
対応して実行される。ワード更新を待たない一
次記憶装置405は、整合性グループに関連し、ス
テップ1150では、現行整合性グループの残っている
レコード更新（最小時間1）を識別する。それらのグル
ープが、次の整合性グループに渡される。それらのグル
ープ内順序番号504は、空バンプで終わり、その操作
時間間隔の間にすべての残取りレコード・セットが読み
取られることを示すに足る。空バンプがない場合
には、現行ソフトウェア整合性グループの最後のレコー
ドを定義するステップ1120を、レコード残取り時間
507および更新時間610と併せて使用して、一次記憶
装置405におけるアプリケーション・リクエスト（入
力）を処理し、更新を完了する。

【0060】ステップ1160は、完全災害復旧の制約（出力がゼロになる）が適用される場合である。

プロセスに入っている整合性グループ全体を最初から再適用することができます。このため、2次のDASD 4 1 1 6の入出力が行われたか、どの入出力が行われなかったか、どの入出力が処理中かなどを追跡せずに、遠隔シヤードレーニングを行うことができます。

【0061】二次入出力書込み

ステップアップ1160の重要な構成要素は、二次側431が一次側421から後れをとらないように、PDM414によってコードが二次DASD416に格納され、主に、線々な二次DASD416への複製の出力操作を同時に実行することである。必要とすれば、二次DASD416を一度に1つずつ連続して書き込むと、二次側431は一次側421からわずかに遅れて戻り始める。単一チェーンの二次コマンドワード(CCW)連鎖を介してデータの二次装置宛の堅固性グループウェアにレコード等を含めれば、二次側431ではさらに高い効率を得られる。それらの単一CCW連鎖内では、そこで行われる二次DASD416の発生順、トラックでの入出力操作が一次ポリシーによって最適化されている限り、その入出力操作をさらに最適化することができ。

【0062】特定の整合性グループ用の二次入力操作を単一C/W域内にて最速で行われる場合、一部は二次入力操作のバタケンに基づいて行われ、一部は二次入力操作の物理特性に基づいて行われる。最速は二次は、二次DASD416がカウンタ/キー/データ (CKD) 方式、拡張カウンタ/キー/データ (EKD) 方式、固定バウク方式 (FBA) などに応じて、多少の異なる操作性がある。その結果、所与の時間隔の間に1つの二次DASD406に対して行われる複数のWRITE出力 (m) は、1つのSTART入出力操作16のポリームに対する単一のSTART入出力操作16に相当するところがある。このように二次制御制御415に対するSTART入出力の回数をm:1に最適化する。二次DASD416は格納をたらすに及び、それに、一次側421のレコード更新をもつ精密にシャドーイング421が行われる。

【0063】正常な送受信データ・シャドールーピングとそれによって一次出力の最上位の重要点は、一貫性のあるエコノミーによる二次出力の最適化である。所与の二つの回路は互いにエラーを最小限にすることである。両方の二次書込みで失敗すると、その後の従属書き込みが条件付けられて同時にエラーを最小限にすることである。両方の二次書込みに伴わずに記録される必要がある（たとえば、更新の際にはデータベース用の最新の更新書き込みが共に記録されている）のでデータベース・レコードが更新されたことを示すログ項目は、二次DSD416のエコノミーの順序を整合性に違反する。

【0064】その更新失敗が復旧されるまで、失敗した二次DASD416のコピーはアプリケーションの復旧

に使用できなくなる。失敗した更新は、SDM414に
よってPDM404から現行コードを要求することで修
正できるはずである。その間、PDM404が現行更新の
整合になり、そのため、PDM404が現行更新の意
識し、それ以前の他のすべての更新がPDM404によ
って処理されるまで使用できない。通常、失敗した更
新の復旧に要する時間は、十分な災害復旧保証のために
受け入れられなければならないほど長い非稼働タイムを示す。

【0065】効果的な二次側431の入出力最適化は、
所与の整合性グループについて書き込まれるデータ・レ

コード・セットを検査し、ECKD方式などの次SDA
SD416の特定の方式の検出に基づいて連鎖化検査す
る出力エラーが発生した場合には警告を出力する。こ
の間に、CVC連鎖を再実行できるように、または、二次
初期プログラム・ロード（IPL）復旧の場合に、デー
タを紛失せず、整合性グループ全体を再運用するよう
に、入川エラーなどの復旧に特化したものがある。

【0066】図12は、ECD方式用のすべてのWRT I T E入出力の組合せに対応するCCW連鎖を要約した完全整合性グループ復旧(F C G R)の図例を要約して示すもので、ここではCCHRRレコード形式が使用される(シリアル番号、ヘッド番号、レコード番号)。図12は、1つの整合性グループの範囲内CDA SDトラックに対して行われるWRT I T E入出力操作の可能な組合せをそれぞれ検定することによって作成される。図12のF C G R規則(図14および図15に記述する)は、整合性グループを適用する際のエラーについで完全復旧を行うためにデータ復旧(二次CDA SD 1

6の入出力並進はCCW連続)を管理する場合に値ともなる。図12に示すFCGR規則は、新しいWRIT E入出力操作が追加されるたびに適切に拡張されることとなる。これらの規則は、二次側431のハードウェアまたはソフトウェアで実装することができ、FCG R規則は、同一DASDトラック分析に対するREAD レコード・セットを、一次DASD406のWRIT E入出力タイプ、検索引数、およびカレント・ワールド とキー・ワールドの検査に還元するので好都合であ

【0067】図12に示すように整合性グループの書き込み操作を検査せずにDASDトラックが書き込まれると、前に書き込まれたデータ・レコードの書き込みができなくなる可能性がある。たとえば、以下の内容の連鎖があると想定する。

レコード5へのWRITE UPDATE
レコード1へのFORMAT WRITE
この場合、レコード1とレコード5は同じDASDトラ
ックに存在し、レコード1がレコード5の前に置かれ
る。レコード5はUPDATE WRITE CCWに
よって更新されるが、FORMAT WRITE入出力

[illegible]

【0088】図16のステップ1410～1470は、図17のステップ1160によって表され、図12に定義されているCGR規則を使用するプロセスの詳細を示すものである。ステップ1410でSDM414は、実行環境におけるプロセスの各種レコードを2つのカテゴリに分類する。第一のカテゴリは、同じ二次DADポリシー向けの入出力命令を含む、第二のカテゴリは、第一のカテゴリに含まれるレコードのうち、同じCCH向けのレコードの入出力命令を含む（すなわち、同じDASDトラックに更新する）レコード。

【0069】 現行整合性グループのレコードへのカテゴリ分け後、ステップ1420では、トラックレコードのデータ配置を識別し、アプリケーションのWRITE出力および/または、アプリケーションのWRITE出力を二次DASD416のWRITE出力に、ECKD方式のFCGR規則（図12を参照）に適合させる。SDM414は、ステップ1430で、同じポリシーに対するWRITEのDWR1430、同じポリシーを単一出力CWCW連鎖にグループ分けする。ステップ1440は、実際の二次DASD416の書き込みの検索引数と特定のレコード・データ（CKDフィールド）に於いて、それぞれの二次DASD416のヘッド・ディスタンス・テンプレ（HDA）を移動させることを含む。

[0070] ステップ1450では、後続の雷込み操作によって、前の雷込み操作またはDASD格納引数（ここでは省略されるレコード番号）を参照するかどうかを判断するために図12のFCGR規則を使用し、どうやら判別するため図12のFCGR規則を使用し、第二のカテゴリ（通常、レコードを区切るトランシットを構成するレコードについてCREATE SECT BUF FERSと2と比較する。FCGR規則に従うと、エクスポートが発生した場合には、一側面421からユーザ更新を再受け取らなくても、SDM414は整合性グループ全体を再書き込みで二度行うことになる。SDM414が現在行なわれるグループは二次DASD416に適用した後、ステップ1460では、状態テーブル（図7）とマスター・ジャーナル（図8）を更新する。

【0071】ステップ1470は、次の整合性グループ（現行整合性グループになるもの）を獲得して、処理を

制御装置ととの通信リンクは変更可能である。また、同期速隔二重化モードの一次側と二次側を分離する際、本発明の指示から逸脱せずに今後の本発明の実施例における改良型の通信リンクによってさらに延長することができる。

【0075】まとめとして、本発明の構成に関して以下の事項を開示する。

【0076】(1) 一次側が一次記憶サブシステムに連結された一次プロセッサを含み、前記一次プロセッサが一次記憶サブシステムへの出力意図は更新の原因となる、少なくとも1つのアプリケーションをそこで実行し、前記一次プロセッサがそこで実行される入出力エントリ・回復プログラム(入出力ERP)をさらに有し、二次側は二次記憶サブシステムに連結された二次プロセッサを含み、前記二次側が出力意図は更新を同期的にシャドウイングし、前記二次記憶サブシステムが入出力ERPと通信する、遠隔データ・シャドウイングを提供する方法において、前記方法が、(a) 前記一次記憶サブシステムから前記二次記憶サブシステムに入出力意図は更新を送るステップと、(b) 前記一次および二次記憶サブシステム間に二重対が確立されているかどうかを判断するステップと、(c) 一次記憶サブシステムから一次プロセッサに障害が発生した二重対を報告するステップと、(d) 二重対の障害が発生した場合に、少なくとも1つのアプリケーションから入出力ERPに一次プロセッサの制御権を移転するステップと、(e) 二重対の障害の原因を判断するステップと、(f) 入出力ERPから二次記憶サブシステムに二重対障害発生メッセージを送るステップを含む方法。

(2) 一重対の状況が設定されるか回復された場合に前記一次記憶サブシステムが更新の書込みを続行すること特徴とする、上記(1)に記載の遠隔データ・シャド・イメージを提供する方法。

(3) 障害が発生した二重対を報告する際に前記一対記憶サブシステムが前記一対プロセスに出入力状況としてチャネル終了/整理終了/整理チェックを報告することによって記憶(c)に含まれることを特徴とする、上記とがステップ(c)に含まれることを特徴とする、上記(2)に記載の遠隔データ・シャドレーイングを提供する方法。

(4) 入出力 E/R P が一たび状態サブシステムにセンズ入出力を出すことがステップ (d) に含まれることを特徴とする、上記 (3) に記載の遠隔データ・シャドーン・グを提供する方法。

(5) 二重対の障害が発生した場合に入出力ERPによるエラー回復を実行し、二重側でデータ保全性を維持するためのステップ (g) をさらに含むことを特徴とする。上記 (4) に記載の遠隔データ・シャドーイングを提供する方法。

(6) エラー回復が正常に完了した後で入出力ERPから少なくとも1つのアプリケーションに制御権を返すた

めのステップ (h) をさらに含むことを特徴とする、上記記 (5) に記載の遠隔データ・シャドードイングを提供する方法。

(7) エラー回復が正常に完了できなかった場合には入力力ERPを介して次側に障害発生同期遠隔コピー・メッセージを送るためのステップ (i) をさらに含むことを特徴とする、上記 (6) に記載の遠隔データ・シャード・インテグを提供する方法。

(8) ステップ (i) によって、一次超サブシステムへの後続の入出力書込みがさらに防止されることを特徴とする、上記 (7) に記載の遠隔データ・シャドーン・グを提供する方法。

(9) 一次プロセスと二次プロセスとの間でエラー回復のための通信が行われることを特徴とする、上記

(5)に記載の遠隔データ・シャドローイングを提供する方法。

(10) 一次記憶サブシステムと二次記憶サブシステムとの間でエラー回復のための通信が行われることを特徴とする、上記(5)に記載の遠隔データ・シャドーイングを提供する方法。

(11) 一次側と二次側を有し、データ・バックアップのために二次側が一次側から定期的な更新を受け取り、一次側が一次側から定期的な更新を受け取るように、二次側が一次側から十分離れた位置に設置される遠隔データ二重化システムにおいて、遠隔データ二重化システムにおいて、遠隔データ二重化システムを生成する少なくとも1つのアプリケーションを実行する、一次側の一次プロセスと、遠隔データ二重化システムを生成するための一次記憶装置から出力更新を受け取り、更新するための一次記憶

サブシステムと、データ・リンクによって一次プロセッサに接続される、二次側の二次プロセッサと、二次プロセッサに接続され、通信リンクを介してさらに二次記憶サブシステムに接続され、書込み出力更新のバックアップのために記憶域ベースの二重冗差を提供し、受け取ったときの特定の同期順に変込み出力更新が書き込まれる、二次側の二次記憶サブシステムと、一次プロセッサ内において、二次記憶サブシステムに接続され、二重冗差の障害が発生した場合にエラー・回復を行うためのエラー・回復手段とを含む遠隔データ二重冗差システム。

(12) エラー回復手段が、一次プロセスと二次側の両方に二重対障害発生モードを報告することを特徴とする。上記(11)に記載の遠隔データ二重化システム。

(13) エラー回復手段が、二重対の障害の原因を判断しながら、一次プロセスで実行中のアプリケーションを静止する。上記(11)に記載の遠隔データ二重化システム。

(14) エラー回復手段が、一次記憶サブシステムにセ
ンス入出力を出すことを特徴とする、上記(13)に記
載の遠隔データ二重化システム。

(15) エラー回復手段がエラー回復を完了し、二重モ

ードが再確立された後、エラー回復手段が一次ブローセッサで実行中のアプリケーションに制御権を返すことを特徴とする、上記(14)に記載の遠隔データ二重化システム。

(16) エラー回復手段がエラー回復を正常に完了できない場合に、エラー回復手段が一次記憶サブシステムの書込み出力更新をさらに防止することを特徴とする、上記(15)に記載の遠隔データ二重化システム。

(17) 一次プロセッサに連結され、エラー回復手段からエラー・メッセージを受け取るための一次コンソールと、二次プロセッサに連結され、エラー回復手段からエラー・メッセージを受け取るための二次コンソールとをさらに含むことを特徴とする、上記(1)に記載の遠隔データ二重化システム。

(18) 一次および二次記憶サブシステムがそれぞれ複数の記憶制御装置を含むことを特徴とする、上記(1)に記憶の遠隔データ二重化システム。

(19) 通信リンクが、複数のエンタープライズ・システム接続 (ESCON) リンクを含むことを特徴とする。上記 (18) に記載の第 2 重化システム。

[illegible]

入出力更新を受け取るための複数の一次記憶制御装置

と、書込み入出力更新を格納するための複数の一次記憶

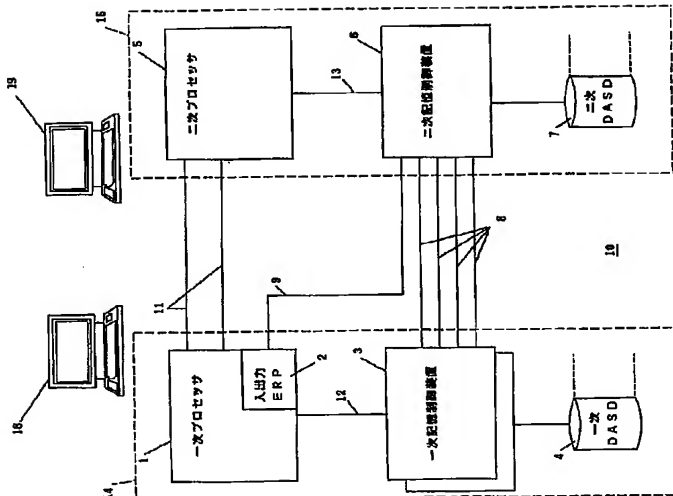
ステップ1410に戻すので、遠隔コピー・プロセスはリアルタイムで続行される。一次側421から二次側431への通信が終了した場合、遠隔コピー・プロセスは、停止する。この通信は、ポリユーモ対がPDM404によってプロセスから削除された場合、一次側が破棄され、また二次側431で特定の引継ぎ処置が行われた場合（災害が発生したとき）、規則的な通信が行われなかった場合、または二次側431で特定の引継ぎ処置が行われた場合に終了することができる。二次側431でジョーナル処理された整合性グループは、引継ぎ操作に二次側4316に適用することができる。一次側421が補足したデータのうち、SDM414によって完全に受け取られていないデータは、発生する。

【0072】要約すると、これまで同明および非同明状態隔データ二重化システムについて説明してきた。同明状態隔データ二重化システムは、記憶域ベースのリアルタイム災害復旧機能を提供する。この選隔データ二重化システムは一次側と二次側を含む。二次側は、データ・シヤードリングのために一次側から書き込み出力更新を同時期的に受け取るが、災害によって一次側が使用不能になった場合に二次側がリアルタイムのデータ可用性を提供できるように、二次側は一次側からデータ・シヤードリングされた位置に配置される。一次側が一次プロセスは複数のアプリケーションをサポートし、この複数のアプリケーションによって書き込み出力更新が行われる。複数の一次記憶制御装置がこの書き込み出力更新を受け取り、書き込み出力更新が複数の一次記憶装置に書き込まれるようにする。一次側の一次プロセスは、データ・シヤードリングによって一次プロセスに連結され、二次側の複数の一次記憶制御装置は、一次プロセスに連結され、二次記憶制御装置に書き込まれている。複数の二次記憶制御装置は、書き込み出力更新をシヤードリングするために記憶域ベースの二重対を提供する。複数の二次記憶装置は、記憶域の一次記憶制御装置から受け取った二次記憶の二次記憶制御装置から書き込み出力更新を受け取る。

【0073】一次プロセス内の出力エラー回復プログラムは、複数の二次記憶制御装置と通信し、二重対称障害が発生した場合は、この出力エラー回復プログラムが前記複数のアプリケーションから制御権を引き継ぎ、エラー回復を行いながら、障害が発生した二重対称特徴に関して一次プロセスと複数の二次記憶制御装置に二重対称障害発生メッセージを出力する。出力エラー回復プログラムは、エラー回復が失敗に終わった場合に一人および二次プロセスに要求を出力しながら、さらに出力要求は更新が行われるのを防止する。

【0074】特に本発明の実施例に関連して本発明を図示し説明してきたが、当業者は、本発明の精神および範囲を逸脱せずに形態および詳細の様々な変更が可能であることに留意されたい。たとえば、一次プロセッサと二次プロセッサとの間、および一次記憶制御装置と二次記憶制御装置との間の接続関係は、図1に示すものと異なる。

【図1】



32

込み更新を二次側にシャードイングするための改良された設計および方法を提供することができる。

【図面の簡単な説明】

【図1】 同期遠隔データ・シャードイング機能を有する災害復旧システムのプロック図である。

【図2】 図1の災害復旧システムにより同期遠隔コピーを提供する方法を示す流れ図である。

【図3】 入出力エラー回復プログラム（入出力ERP）操作の方法を示す流れ図である。

【図4】 非同期遠隔データ・シャードイング機能を有する災害復旧システムのプロック図である。

【図5】 図4の一次側からの読取りレコード・セットの前に付く読頭部ヘッダを示すデータ形式図である。

【図6】 読取りレコード・セットを構成する各種フィールドを示すデータ形式図である。

【図7】 ポリユーーム構成情報を識別する状態テーブルである。

【図8】 図4の二次側が使用するマスタ・ジャーナルである。

【図9】 整合性グループを形成するためのシーケンス例である。

【図10】 整合性グループを形成するために情報および読取りレコード・セットを収集する方法を示す流れ図である。

【図11】 整合性グループを形成する方法を示す流れ図である。

【図12】 DASDトラックに対する所与の入出力操作シーケンスの場合のECKD方式装置用の完全整合性*

【図13】 図12のテーブルで使用している規則の説明の構成を示す図である。

【図14】 図12のテーブルで使用している規則の説明の一部である。

【図15】 図12のテーブルで使用している規則の説明の一部である。

【図16】 完全整合性グループ復旧機能を持つ二次側に読取りレコード・セット・コピーを書き込む方法の流れ図である。

【符号の説明】

1 一次プロセッサ

2 入出力ERP

3 一次記憶装置

4 一次DASD

5 二次プロセッサ

6 二次記憶装置

7 二次DASD

8 対等通信リンク

9 エンタープライズ・システム接続（ESCON）リンク

10 災害復旧システム

11 ホスト間通信リンク

12 チャネル

13 チャネル

14 一次側

15 二次側

【図8】

| |
|---|
| 整合性グループ番号 |
| ジャーナル上の位置 |
| 操作タイム・スタンプ |
| 「ソフトウェア整合性グループ」 別にグループ分けされた 特定のレコード |
| 800 マスタ・ ジャーナル |

【図5】

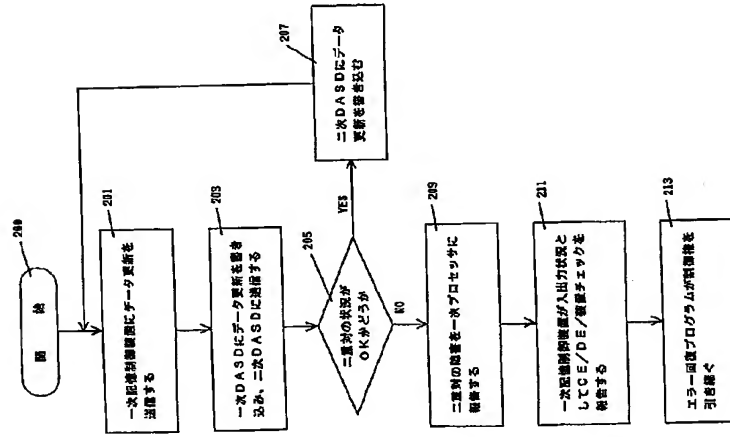


【図13】

| |
|-----|
| 図14 |
| 図15 |

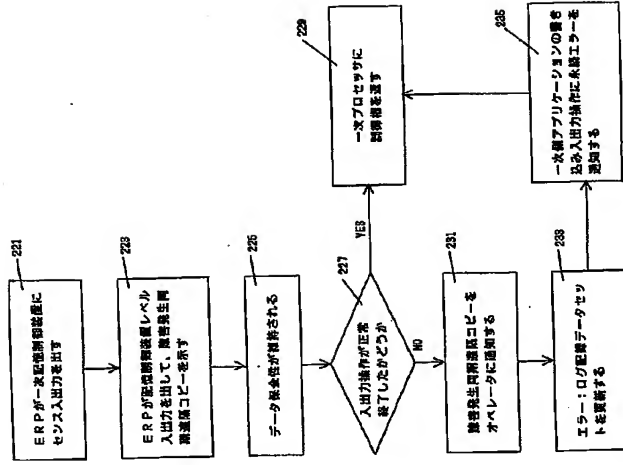
(19)

【図2】

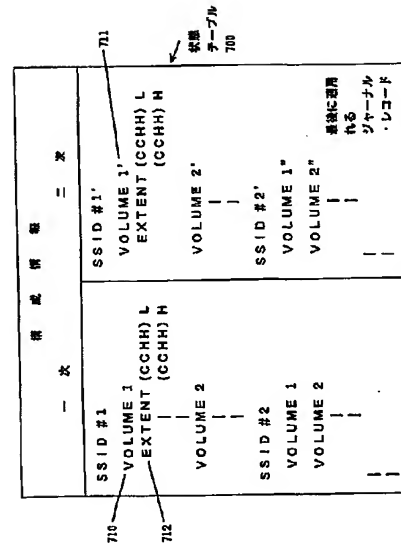


(20)

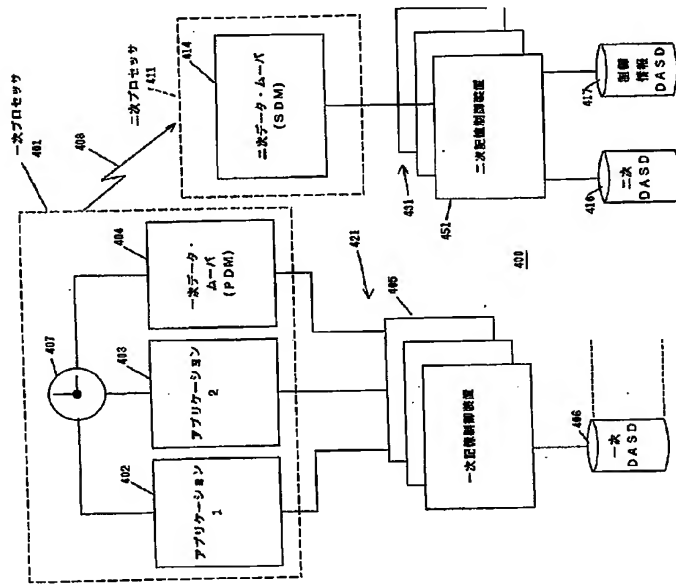
【図3】



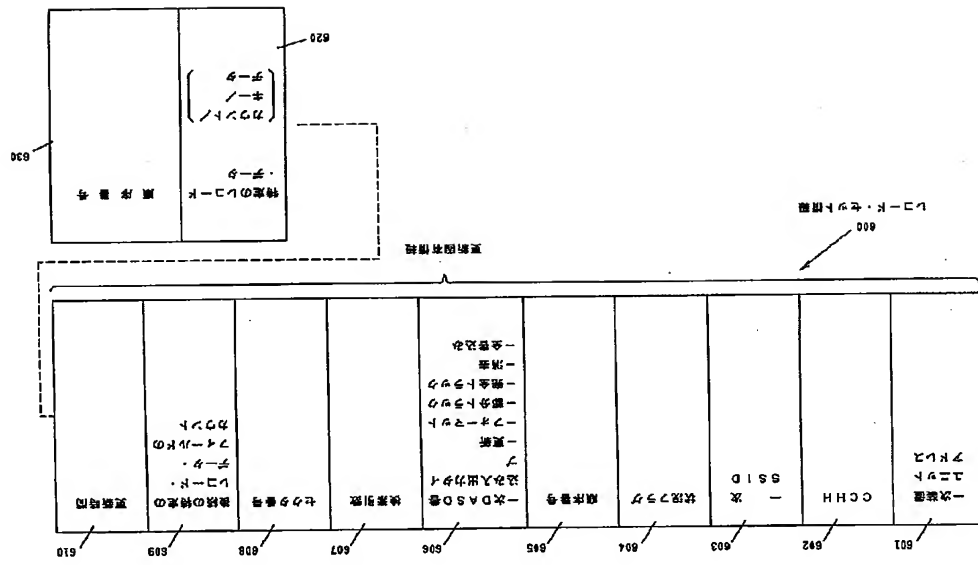
【図7】



【図4】



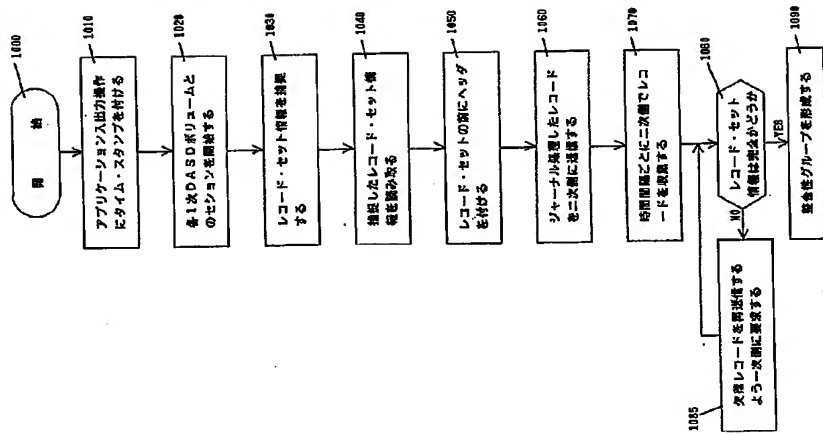
【例6】



【図9】

| 物理記録 装置ID | 操作タイム ・スタンプ | 時間範囲 グループ 番号 | 記録リレコード・セット 更新時間/削除時間 | | |
|--------------|--|--------------------|--------------------------|---------|---------|
| | | | 順序3の1 | 順序3の2 | 順序3の3 |
| SS1D1 | T1 | G1 | 11:59 ㊦ | 12:00 ㊦ | 12:01 ㊦ |
| SS1D2 | T1 | G1 | 12:00 ㊦ | 12:02 ㊦ | |
| SS1D3 | T1 | G1 | 11:58 ㊦ | 11:59 ㊦ | 12:02 ㊦ |
| SS1D1 | T2 | G2 | | | |
| SS1D2 | T2 | G2 | | | |
| SS1D3 | T2 | G2 | | | |
| SS1D3 | T3 | G3 | | | |
| | | | | | |
| 更新性グループ番号1 | | | | | |
| ㊦ 11:58 | 記録リレコードに並べた命令の最も早い操作時間 T1 SS1D全体で最も早い更新時間 | | | | |
| ㊦ 11:59 | | | | | |
| ㊦ 11:59 | | | | | |
| ㊦ 12:00 | SS1D全体での最大更新時間の最小時間 | | | | |
| ㊦ 12:00 | | | | | |
| ㊦ 12:01 | | | | | |
| 更新性グループ番号2 | | | | | |
| ㊦ ---- | | | | | |
| ㊦ ---- | | | | | |

【図10】



【図14】

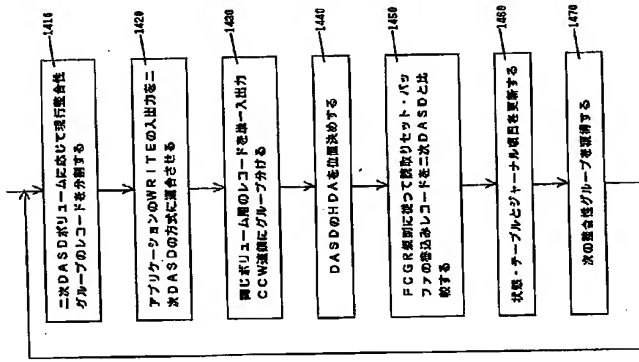
- B-番号1の検索引数より大きい場合、番号1を捨てて両方とも実行する
- C-番号1のレコードが番号2に第1のレコードと等しいかそれ以上である場合、番号1を捨てて両方とも実行する
- D-番号2がR0を返している場合、両方とも実行するかエラーになる
- E-エラー（発生してはならない）
- E'-番号1と番号2が同じレコードである場合、エラーになる
（両側間で消滅をなすを行わずに廃止してはならない）
- F-番号2の第1のレコードがR1である場合、両方とも番号をむか、エラーになる
- G-番号1の検索引数が番号2の検索引数と等しいかそれ以上である場合、番号1を捨ててエラーになる
- H-番号1の検索引数が番号2の最後のレコードより大きい場合、番号1を捨てて、または番号2の検索引数が番号1の最初のレコードより大きい場合、エラーになるが両方とも番号をむ

X-さらに単純化するに於ての手順を実行できず
（番号1の検索引数が番号2の最後のレコードと等しいかそれ以上である）または（番号1の最初のレコードが番号2の最初のレコードと等しいかそれ以上である）しかも（番号2の検索引数が番号1の最初のレコードと等しいか、それ以下である）番号1を捨てて
または（番号2の検索引数が番号1の最後のレコードより大きい）番号1を捨てて
場合
エラーになる
または両方とも番号をむ

【図15】

- J-番号2のレコード（または検索引数）が番号1の最後のレコードより大きい場合、エラーになるが両方とも番号をむ
- K-番号2のレコード（または検索引数）が番号1の検索引数より大きい場合、エラーになるが両方とも番号をむ
- L-番号1の検索引数が番号2の検索引数と等しいかそれ以上である場合、両方とも番号をむかまたは番号1を捨ててエラーになる
- M-（番号1の検索引数が番号2の検索引数と等しいかそれ以上である）場合、番号1を捨てて、または両方とも番号をむ
- N-番号2の検索引数が番号1の最初のレコードより大きい場合、エラーになるが両方とも番号をむ
- R-番号1を捨ててもよい
- T-番号1を捨てなければならぬ
- W-両方とも番号をむ
- W'-番号1と番号2が同じレコードを持つ場合、番号1を捨ててかまたは両方とも実行するかまたはレコードを置き換えて一方の置換を実行する

【図16】



フロントページの続き

- | | | | |
|---------|-----------------------|---------|-----------------------|
| (72)発明者 | ウィリアム・フランク・ミツカ | (72)発明者 | マイケル・アロイス・ボールソン |
| | アメリカ合衆国85718 アリゾナ州ツー | | アメリカ合衆国95037 カリフォルニア州 |
| | ン イースト・ラズベラダ 3921 | | モーガン・ヒル ホウィツバーク・ド |
| (72)発明者 | クラウス・ウィリアム・ミツケルセン | | ライプ2901 |
| | アメリカ合衆国95120 カリフォルニア州 | (72)発明者 | ロバート・ウェズリー・ショムラー |
| | サンノゼ シーナリー・コート・ドライ | | アメリカ合衆国95037 カリフォルニア州 |
| | 6557 | | モーガン・ヒル ピエドメント・コート |
| | | | 17015 |